Grasping cities through literary representations: a mix of qualitative and quantitative approaches to analyze crime novels

Rauscher, Janneke

Grasping Cities through Literary Representations. A Mix of Qualitative and Quantitative Approaches to Analyze Crime Novels

Janneke Rauscher[*]

Abstract: *»Städte durch ihre literarischen Repräsentationen erfassen: Ein Mix aus Qualitativen und Quantitativen Ansätzen zur Analyse von Kriminalromanen«.* Contemporary crime novels often contain detailed literary representations of urban life worlds. These stagings can provide access to city-specific patterns and structures of thought, action and feeling, as well as locally established bodies of knowledge and processes of sense-making. Therefore, their systematic analysis can generate insights into the intrinsic logic of cities. To grasp such patterns on city level a preferably broad empirical basis is needed, but the study of large amounts of literary works poses a methodological challenge. This article presents a mix of methods that permits the analysis of vast quantities of (literary) texts through combining the classical qualitative close reading with elements from computer-aided qualitative content analysis, basic instruments from corpus linguistics and the methodology of distant reading in an iterative research process. It illustrates how to analyze qualitative data also quantitatively and on different levels with regard to social and spatial aspects of the depicted life worlds, thereby showing how novels could be used as data basis for urban sociology and interdisciplinary research questions about the distinctiveness of cities.

Keywords: Intrinsic logic of cities, urban studies, novels as data, quantitative literary analysis, close reading, distant reading, corpus linguistic, mixed methods, space.

1. Exploring Literary Representations of Cities: A Methodological Challenge

Literary representations of cities, especially the often detailed depictions of urban settings and the diverse approaches to narrate the peculiarities of city life that can be observed in fiction, are a fertile ground for further research: they offer a way to investigating attributions to and discourses about cities as well as

[*] Janneke Rauscher, Goethe Universität Frankfurt am Main, Institut für England- und Amerikastudien, Grünebergplatz 1, 60629 Frankfurt am Main, Germany; janneke.rauscher@web.de.

Historical Social Research 39 (2014) 2, 68-102 | © GESIS
DOI: 10.12759/hsr.39.2014.2.68-102

the literary staging of the whole of a city. Due to their manifold potential such literary representations of urban life, urban space and urban spatial structures provide a rich data basis for both literary and urban studies. Approaches from urban sociology that focus on the distinctiveness of cities (in contrast to the long-standing paradigm of a universal urban experience or 'the' city in general), on potentially city-specific aspects, patterns and the unquestioned and often subconsciously operating knowledge of how things are said or done in a specific city, can profit from a study of characteristic and differing modes of representation and staging of one city in contrast to others. Obviously, such an explorative study aiming at the investigation of city-specific structures should not rely on the analysis and comparison of a small sample of literary texts or passages, but needs to be grounded on a preferably wide empirical basis: city-specific patterns should occur across a wide range of individual authors and texts of one city and differ sufficiently in comparison with representations of other cities to show a systematic impact instead of singular peculiarities. Through a quantitative, computer-assisted comparative analysis of a fairly large quantity of crime novels (n_{all}=240) and the cross examination of other kinds of data that relate to the production of these novels as well as their reception, our ongoing project explores the question if and how the literary staging of urban life worlds can be traced back to city-specific factors of influence and reservoirs of pre-reflexive knowledge and attitudes.[1] Besides this mix of different kinds of data, the quantitative analysis of literary texts themselves (being inherently qualitative in nature) poses a methodological challenge. These texts provide a vast data basis for the question of the distinctiveness of cities which has so far remained underexplored in the field of urban studies as methods and procedures how to conduct such a large-scale analysis of literary sources are not readily available. This article presents a mix of methods to show how such a study could be conducted.

Standard procedures of the analysis and interpretation of literary texts are not suited for large-scale investigations across vast amounts of novels. The main method applied is the qualitative *close reading* that can be defined as paying "close attention to textual details with respect to elements such as setting, characterization, point of view, figuration, diction, rhetorical style, tone, rhythm, plot, and allusion" (Rapaport 2011, 4; cf. Brummett 2010). While close reading is vital to interpret individual literary works, it is, not least because of its time-consuming nature, less suitable for the analysis and comparison of

[1] The DFG-funded project "At the scene of crime: the intrinsic logic of cities within the medium of contemporary detective stories", on which this article is based, is part of the LOEWE-Project Group "Intrinsic Logic of Cities" at the Technical University Darmstadt. The other projects within this superordinate project group investigate the intrinsic logic of cities simultaneously within problem discourses, practices and images of city-marketing and economic practices of hairdressers.

large amounts of literary texts. But how should we proceed if the sheer amount of books we want to analyze is simply too high to read them all in detail?

Quantitative and computational approaches aimed at a generalizing analysis of large amounts of texts do exist and have been discussed since at least the 1960s (cf. Hoover 2008). Existing methods and techniques, primarily developed within the domains of corpus linguistics (Hunston 2006) and stylistics (or stylometry, cf. Biber 2011; Mahlberg 2007), focus on very general lexical and/or grammatical language patterns and are used for authorship-attribution (see e.g. the studies of Burrows 1992, 2007; Craig 2004; Fischer-Starcke 2010; Hori 2004; Hoover 2001, 2002), to analyze the style of one author or just one single text in comparison to others of the same period or genre (e.g. Burrows 1987; Craig 1999; McKenna and Antonia 2001; Stubbs 2005), or to explore the possibilities of automatic computational genre-attribution (Allison et al. 2011), to mention just a few examples. The promising possibilities of computational approaches to language, literature and the humanities in general, and especially the wake of a new methodology in literary studies focusing on the "*distant reading"* (Moretti 2000a, 2007, 2013) of quantifiable aspects of literary texts or meta-data (like titles or place and year of publication), provide a starting point in the search for new methods for the analysis of social and spatial aspects, structures and patterns across a large number of novels. Instead of applying quantitative methods to identify very general features in large amounts of data *or* qualitative methods to interpret a few examples in depth, we should connect both to enhance and strengthen our analysis and interpretations in the depth as well as in the breadth. This also can help to lower the relatively high threshold of quantitative analysis, more often than not confronting researchers with 'cryptic' lists of words and numbers whose meaning cannot be accessed easily (cf. Beatie 1979; Corns 1991, 128; Potter 1988), and to overcome the "failure to develop a useful dialectic between computer-based and other methods" (Corns 1991, 128) which seems to prevent or at least aggravate a pragmatic approach to the combination of both methods.

To analyze large quantities of qualitative data on different levels, a mix of methods is needed that enables to move *from close to distant and back again.* This article gives a detailed account of our multi-phase research design and the methods and procedures of each phase, showing how a mixed approach can provide the connection between computer-aided and more traditional methods that is required when analyzing large amounts of novels with regard to social and spatial issues (i.e. cities as inherently spatial phenomena), and, thereby, also how literary texts can be used as a substantial data basis for urban research. Before the peculiarities of this kind of data and the methods are discussed in further detail, we will introduce the sociological context in which our approach was developed.

2. The Intrinsic Logic of Cities

Central to the *Eigenlogik der Städte* (*intrinsic logic of cities*) approach in urban research is the question if and in which ways cities form distinctive structures of thought, action and feeling; specific, unquestioned and often subconsciously operating stocks of knowledge of how things are done resp. how to make sense of *this* city in contrast to others (see especially Berking 2008, 2012; Berking and Löw 2008; Löw 2008a, 2008b, 2012, 2013). According to this view, each city forms its own *urban doxa*, a "classificatory principle that commends a specific view of the world" (Löw 2012, 310; cf. Berking 2008, 24-8; 2012, 320-2) which serves as background to establish and fathom meaning, and to position and assess practices, discourses, images, ways of speaking and narrating as legitimate, intelligible and 'authentic' in or for *this* city, while others are rejected. This worldview has grown over time, forming and being based on a "historical web of meaning" (Berking 2012, 322), the *cumulative texture* of a city (ibid.; cf. Suttles 1984) according to which the tacit assessment of each new 'layer', each representation or discussion, takes place. In other words, every city forms and provides "its own local background and its own version of 'how things are' and 'how things have to be done'" (Berking 2012, 321).

2.1 Literary Representations as Medium

The underlying hypothesis of our project is that this distinctive local background also influences the literary staging of the respective city, and, therefore, that literary representations as data provide new and interesting insights to investigate the intrinsic logic of cities. To substantiate this hypothesis, we systematically analyze and compare the literary representations of four different cities (Birmingham, Glasgow (both UK), Frankfurt on the Main and Dortmund (both Germany)) across a vast amount of contemporary crime novels (n_{all}=240), each of which is set in one of the cities under investigation. The aim is to examine if and how intrinsic structures, place-specific or local bodies of knowledge and distinctive forms of expression can also be traced in the literary texts. Circulating images of a city apparently have to be relatable to the local background knowledge to be viewed as 'authentic' depictions (or to be rejected as such), to give a recognizable account of this city and not another for its readers, at the same time adding to the cumulative texture and this body of knowledge for the city they depict. The literary representations and images inform and build upon each other or come into competition. Also, the ways in which characters speak and 'do things' can (and should, if our hypothesis holds true) be recognizably different from city to city, for the very same reasons: e.g. to be recognizable (and authentic) as a detective from Glasgow, not only a

Glaswegian accent, but a specific kind of banter, a certain wittiness seems necessary – and would appear out of place for an investigator from Birmingham.[2]

It is crucial to mention that the focus of our investigation does not solely lie on the level of the *story*, the 'what' that is told in a narrative, but also, and maybe even more prominently, on the level of *discourse*, referring to all the different aspects of 'how' a story is told.[3] This comprises patterns on the lexical level as well as the arrangement of topics, themes, arguments and their relation and connection to each other within a novel. As already mentioned, if one wants to explore patterns and peculiarities on the level of the individual city, it would not be sufficient to rely on the analysis of a few outstanding novels. This would only produce anecdotal evidence or insights into the peculiar representation of a city by author x or in novel y instead of a systematic empirical analysis of the circulating images and strategies of narrating this city in contrast to others.[4] Therefore, we tried to gather a representative set of novels for all four cities under investigation (see 5).

To put it in a nutshell, we want to find out what all these various crime novels set in one city do have in common and in which ways they differ systematically from those set in other cities. Which locations, places and spatial units are selected and frequently presented, and in which terms? Which narrative, rhetorical and literary devices are used for their depiction? Are the locations described realistically or metaphorically, concretized or abstracted, topographically or topologically? Which discourse strategies and semanticizations, which ways of speaking about the city can be found? And in how far can these commonalities and differences be related to city-specific variables? If the guiding hypothesis about an intrinsic logic of cities holds true, all novels set within one city should have more in common than they have with those novels from other cities. Through a comparison of elements that lie on the micro-level of words, their frequencies and their frequent co-occurrences; on the mid-level of topics, themes, tropes, discourse strategies and cultural attributions; and on the macro-level of the whole of one corpus in comparison to the other corpora, we investigate the possibility of city-specific structures and patterns in their literary staging.

---

[2] Our parallel analysis of Amazon reviews (n_{allrev}=1862) of the crime novels included in our corpora showed that it is indeed the way of speaking which is the main criterion for Glasgow crime novel readers to judge the 'authenticity' and realism of a given novel resp. the way the city is depicted. Interestingly, the criteria or aspects on which readers base such judgments differ from city to city: for Frankfurt, it is a minutely and realistic naming of streets, areas and description of places, for Dortmund the characterization of the figures and allusions to events that happened there in reality, and in the case of Birmingham, the readers virtually never even mention the city or only make very general references about the location of the plot, often simply referring to 'the place the story is set'.

[3] For this distinction between story and discourse in literary studies see especially Genette (1994).

[4] It is indeed such a kind of anecdotal evidence for which novels are very often cited in sociological works (cf. Carlin 2010).

3. Crime Novels as Data

3.1 Possibilities and Advantages

The genre of crime fiction offers some advantages for an exemplary and explorative investigation of potentially city-specific structures in the medium of literary representations:

1) The *sheer quantity of crime novels* available for each of the cities under investigation makes a large scale study and comparison possible in the first place. Moreover, the number of crime novels available also hints at the popularity and impact of this genre – as does the frequent appearance of crime novels on best-seller lists. Crime fiction is one of the best-selling and most widely read genres today.[5] Especially in Germany, the subgenre of *Stadt-* and *Regionalkrimis* (roughly translatable with city crime novels or regional/local crime novels) has grown remarkably over the last decade, with many regions resp. cities starring in their 'own' crime novels.[6] This indicates not only the focus that this genre seems to place on place, but also that the literary staging of cities and the images that are produced within this genre are widely distributed and, therefore, potentially influence the way how these cities are thought of and perceived in the public eye.
2) The *serialized format* that is very often used in crime fiction poses a specific framework for the depiction of the cities in question. Series narrate and depict their diegetic world, i.e. the city they are set in, in an ongoing, cumulative and consecutive way. The image and representations of a city can thus be built over time and should, to avoid irritation on the part of the readers, be consistent or at least legitimate and make plausible varying or even conflicting points of view and conceptual variation. The serial format of many contemporary crime novels thus offers the opportunity to trace consistencies, reiterations and changes in the literary representations in a process-oriented way.
3) The *genre resp. sub-genres* themselves offer conditions that provide a *stabilizing contextual factor for the analysis*. Especially crime fiction is said to rely on strong genre conventions that largely determine or at least influence its form and content, e.g. the structure of the plot, characterization, figura-

[5] According to the Börsenverein (2013) the branch of fiction that is labeled as "suspense" produces the second biggest turnover in Germany (27.9% of the total turn-over of the German book market in 2011), while the "crime thriller" genre is even the most popular in the UK market with "sales approaching £200 million" for 2012 (Membery 2013).

[6] See, for example, the map of 'Regionalkrimis' in Germany as presented in the ZEITMagazin on January, 16 (Stolz 2013).

tion, etc.[7] Many aspects of crime fiction such as its spatial structures, too, have been analyzed as genre-specific conventions (Schmidt 1989; Suerbaum 1984). Following this line of research, the sub-types of the genre have also been examined according to their spatial patterns: whereas the sleuths of mystery novels can be associated with clearly delimited or "locked rooms", the hardboiled, private eye detective operates in an urban setting in which different places and spaces are depicted in short succession (cf. Alewyn 1998), so that the city becomes and "intractable terrain, to be grasped only in a fragmentary way" (Horsley 2005, 71-2). Genre conventions possibly pose a restriction on the potential city-specific patterns in so far as they have a strong structuring influence on the novels under investigation. With regard to these constitutive effects of generic conventions our project asks if and in which ways the distinctiveness of cities can establish influence alongside genre-structures, or in how far the latter are altered and varied by city-specific factors. Generic conventions also provide a backdrop for our interpretation of the different aspects and can possibly supply explanations and interpretations for certain patterns.

4) The *dominance of realistic literary strategies*, including such aspects of social and spatial mimesis as the very detailed description of existing places, buildings, streets and events typical of contemporary crime writing provides rich data also on the content-level. Due to their realistically appealing depiction of characters, events and themes these novels can be viewed as specific forms of space-constituting (and space-constituted) practices of imagination and narration. They are bound, at least to a certain degree, to the life worlds of the specific city they depict, and frequently establish recognition effects through references to the actual world. While this spatial and social realism especially holds true for the German phenomenon of *Stadt-* and *Regio-alkrimis*, this trend can also be witnessed in the UK resp. around the world (cf. Evers 2009).

3.2 Methodological Challenges

Besides the aforementioned advantages of the genre that we chose for our analysis, there are also some methodological challenges that arise from literary texts in general, especially when one wants to employ a computer-aided technique for their analysis. The *literariness* of the texts in question distinguishes this type of data from those that are usually used in urban studies (e.g. interviews, ethnological descriptions, surveys), and is one of the arguments against the application of computer-aided and quantitative methods for the analysis of literary texts in general (cf. Van Peer 1989). Especially the highly figurative

[7] These conventions can be (and are, in fact, frequently) varied, newly combined or used in a transgressive way.

meaning of literary language and the complex interplay of different kinds of context in fiction pose major problems for computational literary analysis.[8] While literary texts generate meaning that is highly context-related and context-dependent in differing ways (and with respect to different, sometimes overlapping, sometimes even contracting contexts), most computational approaches can only deal with a very restricted notion of context as direct co-text in their analysis (Hoover 2008).[9] A similar problem arises for the recognition and analysis of metaphors, metonymies and other tropes that cannot be distinguished from a more 'literal' use of the words in any automated way to date. Additionally, language itself (literary or not) can be highly ambiguous on the level of single words: homonyms (i.e. words with very different meanings in different contexts), homographs (i.e. words that are written in the same way but belong to different word classes and therefore have different meanings), and polysemous words (i.e. one word having multiple meanings) cannot be distinguished without the analysis of their context in all, or at least many, of its different varieties. While an 'ordinary', human reader can easily perceive different kinds of rhetorical devices and potentially concurring meanings embedded, and infer from the context on different levels which meaning is crucial within a specific passage or which sentence builds a metaphor while another one can be taken literally, computers are 'dumb clerks' (to repeat an often heard argument) that are not able to distinguish between different uses of the same string of characters or recognize and trace the meaning of a symbol, trope or reference.

Basically, most approaches in the field of quantitative literary analysis focus on single words, simple frequency counts, or on the automatically retrieved most frequent words of a certain text or corpus (which, unsurprisingly, are very general ones such as *is*, *that*, *and*, etc.). Major methods in this area are the study of keywords (the key-ness of a word can be defined either by its high and/or unusual frequency in a given corpus or by its importance to the research question), word frequencies, co-occurrences, lexical-clusters, collocational and concordance analysis (cf. Biber 2011; Hoover 2008; Mahlberg 2007). Some approaches take into account the direct and limited context of the word in focus to disambiguate its meaning or function, as does the *KWIC* (key-word-in-context) method developed by Philip J. Stone and others (1966) for the purpose of computer-aided content analysis (cf. Fühlau 1982, 132-3; Früh 2011, 12,

[8] Ranging from the social context in which the literary text has been produced to the intertextual context formed of other texts that stand in relation to the text in question, to the context that the whole of a novel (or series) forms for all the passages, messages, themes, symbols and tropes it contains, to the context of the specific communication situation in which an analyzed speech act, message or description is embedded – all these different and yet related kinds of context have potential influence on the production of meaning of a single word, a concept, theme, symbol or the whole of the novel.

[9] The same problem arises in sociological computer-aided content analysis; see e.g. Fühlau (1982, 127-90).

286). The KWIC approach or *concordance analysis* is widely used and acknowledged today, but the context normally regarded in this method only contains a very limited number of words occurring to the left and to the right of the search-term (usually between three to five), not regarding if these co-words occur in successive but different sentences or are part of the same sentence as the search-term. This restricted conception of context as co-text differs from its broad definition (and importance) in literary studies, while the analysis of the broad context (in this sense) poses a problem for any computer-assisted or quantitative approach to date.

To account for these specificities of our data and the given limitations of a computer aided analysis, we developed a multi-phase research design (detailed in the next sections) that places the emphasis on the *qualitative aspects* of the analysis and interpretation carried out by the researcher while *aiding this analysis* and accelerating it via *computational and quantitative techniques* of text-retrieval, visualization and automated lexical analysis, enabling the analysis of large amounts of text. We address the problems of disambiguation, context and meaning through stressing human interpretation, supported and enhanced through a mix of computational tools that address different aspects and integrate differing degrees of context into the (automated) analysis. An adapted variant of qualitative content analysis, especially suitable to investigate all aspects within a middle-range of context, is combined with basic techniques of quantitative literary analysis (word frequencies and searches, co-occurrence and statistical positional concordance analysis with an adjustable and enhanced range of co-text of 10 words to the left and right around the search terms; see 8.) which address a restricted sense of context as direct co-text on sentence or paragraph level and can be carried out by the computer itself. This mix of methods and tools applied allows to explore the narrow, mid-level and wide context of the retrieved words, sentences and passages through an iterative process to connect *a qualitative close reading with quantitative analysis of automatically retrieved passages and their context,* combining different perspectives on the same data and therefore enabling an examination of specific features across the whole of each corpus, comparison between different corpora also on the level of lexical patterns and patterns of semanticization, and, ultimately, testing qualitatively derived hypotheses.

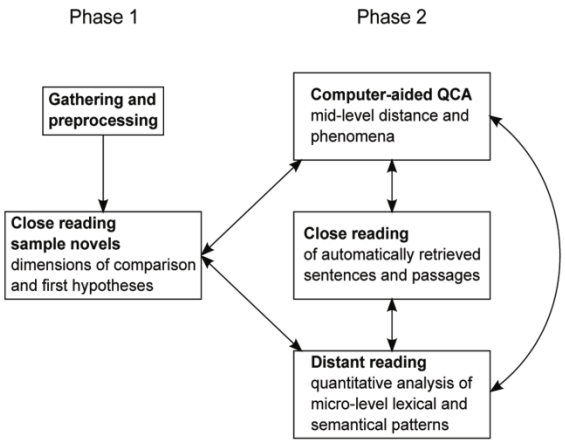
4. The Research Design

Before discussing the different steps and methods of our large-scale analysis of literary texts in detail, a short overview of the general research process of the text analysis (see figure 1) and the gathering and preprocessing of the data is given. This article places the focus on the methodology of textual analysis, and while the project itself comprises also the investigation and analysis of aspects

and data that relate to the sides of production and reception of the analyzed crime novels,[10] these will not be part of the following discussion of our methods. They will only be mentioned in so far as they directly relate to the textual analysis.

The analysis of the literary representations themselves is composed of two phases, whereby each phase consists of different steps. We combine a *sequential explorative strategy* (Creswell 2009, 211-2) with a *concurrent research strategy* (ibid, 213-8), exploring the object of study and generating hypotheses in a *first, qualitative phase* before operationalizing these findings to 'test' them quantitatively throughout the whole of the data in a *second phase,* where the same data is explored in different ways and on different levels. Phase One consists of the preliminary gathering, sighting, choosing and preprocessing of the crime novels (see 5) and the parallel sample analysis of purposefully chosen novels for each city with the basic method of *qualitative close reading*. Based on these first results, hypotheses can be generated and the dimensions of analysis and comparison can be identified that guide the further investigation of the novels and form the basis for the operationalization of these dimensions and the coding scheme (see 6).

Figure 1: Overview of Multi-Phase Research Design



[10] On the side of production, semi-structured qualitative interviews with some of the authors of those novels under investigation from each city are conducted, as well as an institutional analysis of the literary field. The reception of the novels is investigated via the analysis of self-produced statements from the readers of those novels in form of Amazon reviews (see also footnote 2).

Phase Two consists of a *triangulation of quantitative and qualitative methods* (cf. Kelle 2007, 51, 54-7). The combination of different methods of analysis through the use of two different computer tools enables *differing degrees of distance* to the original texts analyzed. This phase can be described as iterative process where the researcher moves between close and distant reading. Based on the results and operationalization that are the outcome of the qualitative first phase, the aim is first and foremost to quantify and refine the findings and 'mine' the whole of each corpus. Therefore, the corpora are approached first in a kind of *mid-level distance* (see 7) through the use of qualitative content analysis (QCA) software (we used *MaxQDA 10*).[11] This makes it possible to trace the impact and relevance of key-terms and first results across the breadth of each corpus, the refinement and enhancement of the qualitative impressions and the analysis and coding of passages in relation to content as well as discourse. Thus it helps to investigate patterns and structures on a mid-level (topics, themes and other aspects in the context of the respective reference which stays embedded in the whole of the text, making close scrutiny easy and allowing the coding and later re-arranging and re-combining of the coded passages in different ways), but also on a macro-level (i.e. the tracing of different topics, themes and spatial aspects throughout all novels in short succession). Eventually, the use of software that is fitted for quantitative literary analysis and distant reading (*CoocViewer*)[12] allows approaching the literary texts from a most *distant perspective* (see 8). This is achieved through an automated analysis across large quantities of texts that aims at the micro-level of the texts (*word frequencies*, their *co-occurrence* and *statistical, positional concordance patterns*), while the visualization of this analysis (i.e. network graphs that display the word-class, absolute and relational frequencies, and the relations between search terms, their co-occurrences and concordances) poses the basis for further interpretation of the results. This distant way of looking at our novels helps to investigate and compare patterns on the micro-level of words, revealing ways of speaking, semanticizations and cultural attributions and enhancing the comparability of the same or similar aspects (words or themes) between the city-specific corpora.

The interplay and flexible handling of the two programs enables the macro-scale investigation of mid- and micro-level patterns as well as a close scrutiny to the contexts in which these patterns appear. This connection of qualitative

[11] We did not adopt the method of sociological content analysis, but only used the software created for this and adapted the instrument of 'coding' for our purposes (see 7).

[12] This tool has been developed especially for our needs and purposes and I want to thank Leonard Swiezinski, Martin Riedl and Chris Biemann from the Department of Computer Science of the Technical University of Darmstadt for the programming of this tool, help with the pre-processing of the data, and the very helpful cooperation with our project. CoocViewer is available as freeware tool and can be downloaded from <http://sourceforge.net/projects/coocviewer/>.

and quantitative methods for the same set of data helps to complement each perspective, providing insights into the structures and patterns of a corpus (and the possibility to compare them) that could not be reached by any one of these methods alone.

5. Getting Started: Gathering and Preprocessing the Data

Preliminary decisions and limitations regarding the question which texts should be included have to be based on the aim of the research. In our study, we focused on the following factors and reasoning:

1) *Setting*: The respective city should be the main setting of the novels.
2) *Genre*: Crime novels in any variety (e.g. detective novel, police procedural, hard-boiled) where included.
3) *Contemporariness:* Operationalized as a 40-year time span (from 1970 to 2011). Each novel in the corpus should have been published for the first time within this period and the main plot should be set within this time span.

About 300 crime novels have been sighted initially, and 240 of them have been scanned and prepared for deeper analysis. This total amount of n_{all}=240 is distributed across the four cities in the following way: Frankfurt provides the largest corpus (n_F=79); followed by Glasgow (n_G=61); Dortmund (n_D=59) and Birmingham (n_B=41) (see also table 1). This uneven distribution is mainly due to the size of the actual 'field' for each city (e.g. Birmingham's 41 novels are all crime novels set in this city that existed until 2011). Additionally, some of the originally gathered novels were ruled out from further analysis according to the criteria stated above, while others had to be excluded because their format posed great problems for the electronic preparation (e.g. comic books). Additionally, we deliberately limited the size of the largest corpora (Frankfurt and Glasgow) through reducing the amount of those novels that belong to very long, ongoing and open-ended series (i.e. with more than nine parts). Here, we limited the analysis to the first and the last novel and three from 'the middle' that were chosen randomly from each of the series in question.

The underlying pragmatic reason that makes limiting the pre-processing and preparation for further computer-assisted analysis necessary is the very time-consuming and labor-intensive nature of these steps. Scanning the novels and transforming them into machine readable text through *OCR* (Optical Character Recognition) makes extensive proof reading necessary, and took more than a year in our case.[13] We used *ABBYFinereader 10 professional*, an OCR-software

[13] At this point I want to thank the *Universitäts- und Landesbibliothek Darmstadt* (University and State Library Darmstadt), their team from the integrated bookbindery for their help with the preparation of the novels, and especially Roland Roth-Steiner from its Center for

that generates tolerable but not perfect results due to the many differing typefaces of the novels. We stored back-up copies (in different formats) of the electronic versions and created four different databases for the content analysis with MaxQDA 10. One novel always is stored as one document, including the metadata for each novel (author, title, original page numbers, chapter-headings, etc.).

To produce a more 'distant' reading resp. a more sophisticated form of automatic computational analysis with CoocViewer, the electronic versions of the novels have to be prepared additionally according to standard practices and basic principles from the field of corpus linguistics. This comprises techniques of parsing, tokenizing and part-of-speech tagging (for a detailed account of these steps see Rauscher et al. 2013, 61-3). The data are also indexed by document, sentence and paragraph to enable the distinction and analysis of co-occurrences on sentence and paragraph level. The following table (table 1) gives an overview of the quantitative characteristics of the different corpora for each city.

Table 1: Quantitative Characteristics of the Corpora (cf. Rauscher et al. 2013, 65)

City	Novels	Tokens	Sentences	Paragraphs
Birmingham	41	4.8 million	336 thousand	142 thousand
Glasgow	61	7.7 million	496 thousand	222 thousand
Dortmund	59	5.0 million	361 thousand	127 thousand
Frankfurt	79	8.0 million	546 thousand	230 thousand

6. Qualitative Close Reading

Considering which sample novels should be read during the initial stage depends on the concrete research interest and should aim at a comprehensive first impression. In our case, we chose the sample according to the statements retrieved from the Amazon reviews: we focused on those novels that readers frequently described as 'authentic' or realistic depictions of the city in question. The second consideration that guided our test-sampling was that as many authors and/or series of one corpus as possible should be included to get a preferably wide-ranging first impression. The application of the qualitative close reading during this early stage was guided by our project outlines and basic questions about the depiction and description of the city and aspects that related to city life and space as described above (see 2), as well as some basic considerations from literary studies (narrative situation, characterization, figuration, form, structure and rhetoric devices). The results for each novel should be

Digitalization (DIZ) for his support of our project with his expertise on this field and making it possible for us to use their scanning infrastructure (which made our project design possible in the first place). Also, I want to thank our student assistants Elena Gontscharow, Daniela Langbein and Marvin Kolb who mainly undertook this preprocessing.

noted, and through a comparison and synopsis of these first results the different dimensions relevant for further analysis can be developed. We established four dimensions subdivided into different aspects and operationalized with certain guiding questions to build a foundation to systematize the further investigation and comparison of the corpora. The following overview specifies the dimensions and offers exemplary guiding questions for their investigation:

1) *Temporal dimension:* Includes patterns and strategies of a) organizing and relating the past, present and future; b) the details, stories or aspects of the (factual) past or history of a city. Guiding questions: Which aspects of the past and the present are narrated, and which role do they have for the plot, the characters, etc.? Which role is ascribed to them for the city? How are past, present and future related to each other? Which narrative patterns are employed to link story and history?
2) *Dimension of designation (spatial)*: Comprises spatial entities that are denoted or referred to as well as the patterns and strategies of naming a) the city itself, b) the districts within the city, c) other cities to which the respective city is compared to or contrasted with directly or to which other relations are established, d) the region (or other spatial units like the nation or Europe) that the respective city is brought in connection with (e.g. as being part of). These different aspects can be either referred to via *denomination of the elements according to the real-world* or via more *descriptive or paraphrasing strategies*. Also, synonymies or nicknames for the cities or the other spatial entities are relevant (e.g. 'Bierstadt' for the city of Dortmund or 'second city' for Birmingham). Guiding questions: How and how often is the city referred to? With which names? Which descriptions, which semanticizations and cultural attributions evoke the city as a whole or its different parts? Which parts of a city are frequently featured, which role do they play in the novels and how are they set into relation to each other and to the city as whole? What is told about other cities and how? To which cities and spatial units is the city in question related?
3) *Social dimension*: Comprises aspects of a) the topics and themes within a city that re-occur, b) those aspects and events that are presented as 'typical' or 'every-day', c) characters, protagonists and actor-constellations that are presented as typical or a-typical, d) local specificities like food (e.g. 'Grüne Soße' (green sauce) in Frankfurt), drink or others. Guiding questions: Which events are presented as such, and which are presented as 'typical', everyday, or only mentioned in passing? Which characters and actor-constellations can be frequently found? How are inhabitants depicted? Does accent or dialect play a role, how do the characters speak? Which attributions are made to inhabitants and which to 'outsiders'?
4) *Architectonic/material spatial dimension*: All aspects of the material space of a city (factual and fictional sites and locations, signature buildings or landmarks, architectural specificities, concrete places and streets) that are

recurrently named or described. Here, the focus is placed on the frequent functions these specific places and their descriptions have for the plot (as direction-marker for movement, as background 'pictures' for the setting, as place of action, etc.) and the patterns that are formed by addressing or discussing this concrete spatial dimension of the cities in the literary discourse. Guiding questions: Which places and spaces are described in which ways? How is the city-space depicted? Do the plot-relevant settings exist in reality or are they purely fictional? Is the material space addressed directly?

The four dimensions have been identified as those which seem to be most directly linked to the literary staging of cities in the genre of (contemporary) crime fiction. To compare the novels we have selected an adaption of George and Bennett's "*structured focused comparison*" (2005, 67), originally developed for the social sciences. It is

> 'structured' in that the researcher writes general questions that reflect the research objective and that these questions are asked of each case under study to guide and standardize data collection, thereby making systematic comparison and cumulation of the findings of the cases possible. The method is 'focused' in that it deals only with certain aspects of the […] cases examined (ibid.).

This allows limiting the analysis to those aspects and passages that are of interest with respect to the city in question and comparing the novels within one city-specific corpus as well as the cities themselves on a standardized basis. Furthermore, it supports a standardized way of literary analysis as teamwork, because each individual researcher poses the same questions, enhancing the comparability of the results.

The dimensions are not limited to the manifest contents or elements that relate to them but always include a focus on how these contents are formally presented and set into relation to each other. Also the semanticizations of the different aspects need to be investigated. The aim is to identify patterns that, on the one hand, emerge through the comparison of novels from different authors and series of a city, while on the other hand differ between the city-specific corpora. We found, for example, that the topic of "traffic" plays a role for the literary representations of each of the cities, but differs across the cities according to the frequency with which it occurs, and, more significantly, according to its semanticizations, cultural attributions and surrounding thematic discourses (e.g., in Frankfurt the passages and statements that relate to 'traffic' are connected to the inner city and the mentioning of commuters from other cities who 'clog' the streets for the residents, representing the heavy traffic as problem that results from the appeal and economic power of the city, while in Birmingham the traffic poses a constant and ordinary problem for everybody and the whole of the city, without the discrimination of groups who 'cause' the problem from those who 'suffer' from it or further reasoning why this could be the case).

The qualitative first phase also revealed that the systematic differences we observed do not seem to lie primarily within basic literary dimensions: neither

could we find city-specific plot-structures (this aspect seems to be more related to the specific sub-genres of crime fiction, e.g. police procedural, detective or mystery novel, or the 'private eye' or hardboiled novel) nor did we encounter city-specific and recurrent ways of narrating or focalizing (e.g. multi-perspectivity or kaleidoscopic accounts). Also, the traditional tropes or metaphors for the city or city-life (e.g. the city as jungle, labyrinth or Babylon) did not so far play an important (i.e. repeated) role for any of the cities in question; neither are there any other that play a systematic role across different authors: e.g. for Birmingham, John Dalton uses the 'city as jungle' metaphor: "Outside, expressway traffic streaked the night with red and amber. He looked at the streams of light – *could be transient firebugs flitting through tall tower moulds of ants. It's a jungle out there after all*." (2002, 43; emphasis in the original). None of the other Birmingham authors present similar or comparable passages.

While the qualitative reading of sample novels already suggested that the city could make a difference within the realm of its literary representations (and not only on the content or story level, but also in different aspects of discourse), the mining of the whole of each corpus is needed to quantify this results and to compare them to the 'unread' of the rest of our corpora to further explore, verify or relativize the initial results.

7. Mid-Level Distance: Quantifying and Testing First Results

The mid-level step between close and distant reading is basically an adaption of sociological, computer-based qualitative content analysis (QCA), which today is often combined with different forms of quantitative resp. quantifying methods (see e.g. Früh 2011, 286-93; Kuckartz et al. 2007; Kuckartz 2010). The use of QCA software (in our case MaxQDA 10) helps first and foremost to code the data and to retrieve passages of interest via simple word searches. Additionally, some aspects of the coding-process (i.e. the allocation of the passages to the dimensions they belong to) can be carried out (semi-)automatically.[14] This step focuses on those passages corresponding to the guiding dimensions and allowing automatic retrieval. To find all the passages that relate to a certain dimension the possibility of computer-based word searches to mine the whole of a corpus at once and quantify the results is especially useful, but makes deliberate operationalization necessary. We encountered the full complexity of the data during the phase of the operationalization of the dimensions, as we could

[14] These are only semi-automated because although the program principally finds all instances of a certain lexical-string, it is required to read through each passage to verify or falsify each occurrence of the word sought (especially when wildcards are used) before the program can "autocode" the verified instances.

not assume that each of the novels in question might use the same terms or spelling for the same topics, aspects or themes (whereas the dimension of designation poses only a minor problem). Some novels include spellings that relate to a certain dialect (especially in the case of Glasgow where many novels feature Gaelic and Scots) or sociolect, others only indirectly refer to topics, places or themes (this is part of the problem of literariness as detailed above). Another problem derives not from the spelling of the text of the novels as such but from their hyphenation as set by the publisher. This challenge can be answered with the compilation of very extensive and principally open-ended word lists to operationalize the coding and quantifying strategies, and to keep track of which words had already been searched.

7.1 Strategies for Operationalizing the Unknown

The operationalization of the terms relevant for each dimension is based on words retrieved during the first phase from the texts themselves. Additionally, it is useful to include terms that complement them: synonyms and related terms from the same semantic field, and also words and names from the factual city. For example, the lists guiding the quantifying investigations in the spatially focused dimension of denomination do not only include general terms (e.g. 'district', 'part') and those parts of the city that we found in the first sample of novels. They also comprise the names of all city-districts to find out which ones are frequently used as setting or reference and how they are connoted and set in relation to each other and the whole of the respective city. Furthermore, some of the aspects can be operationalized usefully with the wildcard function of the software, e.g. in the case of Frankfurt, were many names of the districts end with the suffix -end (Nordend, Westend, Ostend) or -heim (e.g. Ginnheim, Griesheim, Bornheim), so that a search for the suffixes and ensuing sorting and coding of the different instances is possible (which is, at least for some words, a time-consuming method and makes extensive proof reading of the results and manual sorting necessary). The operationalization and investigation should be kept open for new topics, themes or terms that occur during the analysis in the texts themselves.

7.2 Coding as a Way of Organizing and Analyzing

The search for words or word fragments is the point of departure for the further process of coding. It should be noted that our approach does not adhere to the standard procedures for these steps as developed for sociological content analysis. Due to the specificity of the data, the basically explorative aim and open research question of the study, and the simultaneous interest in aspects of content and discourse, the operationalization of all categories and sub-categories beforehand or in an exclusive and fixed way is not desirable (cf. Franzosi 2010). Most passages and examples of seemingly city-specific aspects and

messages relate to more than one category or dimension, making the establishing of mutually exclusive categories for the codes unobtainable. In addition to this, the possibly city-specific connectedness of different dimensions and aspects is of interest and would be lost when assigning one passage to one category only. Besides, there is also a pragmatic reason: we use the software primarily as a means of handling, retrieving and re-ordering the vast amounts of texts according to different aspects. In this pragmatic context the assignment of codes to the analyzed segments offers the best solution to store results of the analysis together with the data and the basis for further steps of analysis with the QCA-software.

Our codes (besides those relating to the uppermost-level of the four dimensions) have been developed during a spirally evaluation process. They are basically comprised of three groups: codes that relate to the content on different levels (e.g. on the micro-level, each search term forms the basis for two codes: one assigned to the word and one assigned to next higher meaningful unit of context with variable length to store and retrieve those searches and passages at a later point; other mid-level codes relate to the topics or themes a passage contains); codes that provide information about the discourse level, i.e. the narrative strategies, functions and structures that can be identified in this passage; and codes that are assigned on a meta-level according to different social dimensions or aspects on the textual level for which the segment could be seen as exemplary (e.g. the code 'difference' is assigned to those paragraphs where the diversity of the described society or certain categories of 'difference' were included, as well as those instances where the relations and social interaction between people with diverse backgrounds are featured).

It is one of the advantages of adapting QCA software and the instrument of coding for the exploration (or mining) of vast amounts of literary texts that the program displays each instance of a word within its original wide context in the novel without limiting or shortening. This makes it possible to read the passages that surround the search terms on screen and to qualitatively decide how much context is relevant for the analysis and understanding of this instance; and allows a connection between the close and the distant levels of analysis, whereby 'distant' here refers to the quantifying as such as well as to the reading of only certain passages, not the whole of each novel. Close reading of specific, automatically retrieved passages helps to get a very good overview of the material in a comparatively short amount of time (compared with reading all the novels from cover to cover), and to directly compare passages from different authors and texts in short succession. It contributes to the understanding of the data and research object and enables a comparison of high quantities of text which would simply not be possible with conventional methods of literary analysis. Also, the different techniques and features QCA tools provide for displaying and converting the codes and their interrelation offer certain advantages for the cross-examination of passages, and help to come to terms with

questions about the distribution of patterns across authors (and, therefore, with the question about the possibility of city-specific structures as such): the interconnectedness of codes (and therefore dimensions and/or aspects of discourse) can be revealed, and displaying the distribution of codes and coded segments across the different documents of one corpus gives an instant impression of how widespread the related discourse or thematic aspect really is, and also which texts or authors are widely excluded. The latter can then be singled out for a next close reading phase to investigate the reasons for their seemingly deviation or variation from the patterns, thereby also helping to further refine the search criteria.

7.3 First Empirical Observations and Why More Distance is Desirable

The use of QCA software helps to transform first, singular and peculiar insights into empirically broad based observations about the specificities and patterns of the literary representation of each investigated city. The following section gives two short examples of our research to illustrate the kind of 'thick' description of patterns and characteristic possible on the basis of this step.

In the case of Frankfurt we found a strong affinity across all authors towards (spatial) mimesis of the real world: many districts, streets and factual details of the city-space are named and very often minutely described; real historical events not only provide the roots for (fictional) plots and stories (e.g. the protests at the end of the 1960s and the beginning of the 1970s around the 'Startbahn-West' (the airport-expansion)), but also very often the history of a certain place or district (the time it has been built, the architect that was responsible, the way it came to look like today) is given, even when protagonists just move along these places or on the streets. Such realism is often connected with a discursive strategy that links the status quo of the city (and its materiality and spatial development) to its past. It is frequently suggested that the city came to be what it is today in the form of a coherent narrative. The corpus of crime novels for Frankfurt resembles sometimes a guidebook to the city, where the real-world references and allusions take a prominent role to reach an effect of realism and authenticity.[15] The city itself features prominently in its literary representations (as setting and topic), and the authors all try to give a 'balanced' description of all the different aspects, events and peculiarities of this city. Most authors feature different parts and aspects in different parts of their

[15] This could also be interpreted as an attempt to give readers who are not familiar with the city a stronger 'sense of place'. Interestingly, those readers that stated in their reviews to be foreigners to the city frequently complained about exactly this strategy (and the repeated use of the city's name), while those that claimed to live in or near Frankfurt recommended it as 'authentic' and 'true' account to their home place.

respective series, giving a very broad picture of the city. The impression of a 'whole' of the city is established through representing them not as completely different life-worlds but all as providing only an aspect that is included into Frankfurt, therefore connecting the differing parts with each other and back to the 'whole'. This discursive strategy and the thematic discourse surrounding certain recurring topics and themes form systematic patterns across all instances. The city is represented as a 'coherent whole' or life-world, where all the different and sometimes contrasting aspects (e.g. between the middle-class detectives and the homeless on the streets which the former very often know by name) seem to have 'their place' and form this social and spatial entity above all through their interrelation. This is tied to the complementary strategy to state that the 'whole' of the city is perceivable in all its parts.

In contrast to this, the representations of Birmingham seem to relate more to the parts of resp. different milieus in the city (especially its 'posh' and wealthy suburban areas to the south and west of the city-center) than to the 'whole' of it. Each series depicts Birmingham as different life-world relating to the districts where the protagonists live, work, and move frequently, and to the milieu the protagonist belongs to (in Frankfurt, the mixing of the different milieus also in the private life of the protagonists is frequently described and reflected). These different worlds are not systematically connected to a 'whole' of Birmingham; they tend to be represented as being not connected with each other, presenting a mosaic picture. This is reflected in the distribution of the single codes which frequently occur across the different novels of one author and seldom across those of different authors. Names of places or streets are comparatively seldom given, and the concrete settings of events (especially the crime scenes) are often fictional or rather described and circumscribed than named. Real historical events or facts are very seldom part of the narratives, and are mostly limited to police- and crime-related aspects (e.g. the famous bombings of Birmingham pubs in 1974 or the case of the 'Birmingham Six'). But while this city seems not to play a prominent role in its literary representations, the quality of its concrete materiality and architecture, especially that of its central area, is often object of reflection and discussions between characters. These discussions center on a notion of frequent or even constant change (of the city and its material dimension) and establish a specific kind of spatial-temporal link between the materiality and the social dimension of the city, e.g. through recurring descriptions and references to 'decaying' places and factories or their conversion (as symbol for the decline of the city's industry, and, again, the frequent change). One of the patterns and structures of discourse about the city that can be found across different texts is this connection of the temporal and spatial dimensions, which centers around a desire for a (better) future with a simultaneous emphasize and recognition of past and present change, presenting the city as being in constant flux.

The use of QCA software helps to trace such relationships between different dimensions and the aspects, topics and discourse strategies that frequently recur in the novels, thereby adding in a substantial way to our research question and making a kind of 'thick' description or characterization possible. But the investigation of semanticizations, lexical structures and the concrete quantification of co-occurrences pose problems with this kind of software. The sorting, coding and the actual analysis of the different passages are still highly dependent on the individual researcher, and the manual counting of all co-occurring words is not feasible without a high risk of inaccuracy or within a reasonable timeframe. Although computer aided QCA-analysis gives a rough overview of thematic structures and first insights into patterns of discourse, it does not provide a possibility to recognize, let alone quantify, co-occurrences.[16] To investigate lexical and semantic patterns that lie on the micro-level of words and their frequent combination, and to answer questions about differing, possibly city-specific ways of speaking in and about the cities or a city-specific stylistic, another step into the 'distance' is needed.

8. Distant Reading

The nascent methodology that Franco Moretti termed "distant reading" (2000a, 56) provides not a clear and consistent method but an umbrella term for diverse methods and approaches. It has stirred some debates in the literary studies community (e.g. Batuman 2006; Prendergast 2005; for an overview of the critique of Moretti's approaches, see Khadem 2012; Serlen 2010) and has been adopted and adapted by others in manifold ways (e.g. Clement 2008; Hayles 2013; Liddle 2012; Mitrić 2007). Distant reading comprises at least two different types of 'distance' to the original texts, as Khadem (2012) points out: An "epistemological distant reading" (ibid., 415) where the particular texts themselves only play a minor role (e.g. as title, as 'instance' of a genre; cf. Moretti 2000a, 2007, 3-34, 2009) and the researcher focuses on large-scale patterns of change over time in the literary field or within the boundaries of specific genres, and an "anatomically distant reading", where "the scholar intends to stay away from an analytical study of the work as a whole in order to, first, gain a definite perspective over it when situated among a huge gamut of other works, and second, be able to trace a limited number of formal elements in the entire

[16] MaxQDA 10 is not working very accurately and seems simply unable to cope with such large amounts of data. Repeated searches for the same terms often produce different frequency counts (and include or exclude different instances and passages), making it necessary to check and cross check large parts of the work time and again. Additionally, the program often functions only on an unreliable basis, stopping in the mid of a session so that one has to begin anew.

range of them" (Khadem 2012, 415; cf. Moretti 2000b, 2007, 35-93, 2011). It is this latter type which seems to be especially prolific for explorative and inductive research. In both cases the tracing of patterns and structures that lie beyond single novels is based on "a process of deliberate reduction and abstraction", where "you reduce the text to a few elements and abstract them from the narrative flow" (Moretti 2007, 1, 53). It allows the researcher to "focus on units that are much smaller or much larger than the text: devices, themes, tropes – or genres and systems" (Moretti 2000a, 57). Or, as in our case: words, themes, discourse structures and patterns of semanticization that reveal possibly city-specific speech and stylistic patterns as well as cultural attributions to topics and the city as spatial and social entity.

Central to both perspectives in this methodology, where distance "is not an obstacle, but *a specific form of knowledge*" (Moretti 2007, 1, emphasis in the original), is the visualization of the abstracted elements. And indeed, *visualization as method* (or instrument) not only achieves 'distance' to the original textual data, it bears advantages and reveals information that can otherwise not be obtained. The presentation of the data and results on a visual basis seems well-suited to alleviate the reservations of the literary studies community against the afore-mentioned 'cryptic' lists of words and numbers resulting from quantitative approaches to literature (see 1). Visualization makes it easier to literally 'see' things, not only for the audience but also for the researcher: it reveals patterns that lie across vast amounts of data and which would otherwise go unrecognized. Through the transformation of textual data into another, visual state of aggregation (as graphs, maps, trees or charts) the researcher is able to perceive connections and relationships between those abstracted elements, therefore gaining another basis for synthesis and interpretation that would not be available otherwise.

8.1 Graphs: Analyzing Micro-Level Phenomena on a Macro-Scale

Distant reading, coupled with basic techniques from quantitative literary analysis, makes the analysis and quantification of patterns on the micro-level of words and sentences (but on the macro-scale of the whole of each corpus) possible and enhances the analysis of potentially city-specific patterns on this level. It helps to test, extend and correct hypotheses that were formed during the previous steps. In our adaption, a program (CoocViewer) is used that combines a word search function, frequency counts and automated computational analysis and visualization of co-occurrences and statistical positional concordances.[17] The parameters of analysis can be adapted to the respective research

[17] Statistical positional concordance analysis considers and displays those words that occur frequently (with a frequency threshold of two) around a search term within a range of -10 to the left to +10 to the right of the term. Their analysis and ordering takes place according

interest and comprise the number of instances (up to the top-30 co-occurrences), choice of word classes considered in the respective analysis (e.g. searching for the top 15 adjectives surrounding the name of a specific part of a city), and the choice of its basis (sentence or paragraph level). The exact quantification of search terms, co-occurrences and statistical positional concordances is possible in absolute and relative frequencies (i.e. as parts-per-million (ppm), a standardized unit of measure indicating how many times the word occurs within one million words of the respective corpus) to enable comparison between corpora of different sizes.

The operationalization of the dimensions resulting from the first phase of qualitative reading and forming the basis for the QCA adaption in the previous step also provides the starting point for the distant reading. After the researcher decided on a search term, CoocViewer automatically retrieves all its frequently co-occurring words and transforms the analysis into network graphs. The co-occurring words form the nodes (or vertices), with the search term always at the center of the network, and the edges between them are weighted according to the strength of connection (based on the frequency and significance of the co-occurrence on sentence or paragraph level).[18] Additionally, each word is colored in the original graphs according to its word class (part-of-speech), displaying grammatical information easily on the same level. This enables also lay-people in the field of quantitative literary analysis to see and analyze connections and relations between co-occurring words (and their interconnectedness among each other) 'on a glance', turning visualization into the starting point for further investigation and interpretation. Moreover, the graphs are interactive and give the possibility to additionally display the concrete sentences and passages within which the computer had found them. This systematically integrates a form of close reading as method for interpretation, which distant reading is lacking so far.[19] The respective references, i.e. the limited co-text of co-occurrences, can be taken under close scrutiny to come to terms with the functions of those elements in their immediate context, the disambiguation of their meaning and a quick way for deeper analysis of semanticization and discourse. They also provide information about the source (author, book, page number) which enables a feedback loop to the close reading and/or the mid-level step, therefore adding the possibility of investigating the wider context in order to enhance one's interpretation. Similar as during the computer-aided QCA analy-

to their frequency and significance of co-occurrence and their recurring position in relation to the search term within this range.

[18] The log-likelihood test (Dunning 1993) is employed to determine the significance of the co-occurrence of two words, with a significance threshold of 3.84 (corresponding to 5% error probability) and a frequency minimum of two occurrences (see also Rauscher et al. 2013, 62).

[19] As Serlen (2010) notes, the actual method of interpretation of the data is "curiously elided in [Moretti's] descriptions of distant reading as a method [...] he conflates the interpretation of data with the representation of data" (219-20).

sis, this also helps to determine whether the diction might be series or author-specific or if it forms city-specific patterns that run across different authors.[20]

Thereby this last step brings together the strengths of more classical approaches to quantitative literary analysis, Moretti's distant reading, and close reading as a method of interpretation. It supports the research in at least three ways: it makes an exact quantification of search terms and co-occurrences possible, thus enabling to answer questions about stylistic and speech patterns; it gives a complementary perspective on the data; and it enables a further study of aspects, topics and terms arising from the first two steps across the breadth of the whole of each corpus, as well as an easy comparison between corpora of different sizes for the same or similar terms or topics.

After this general description the following section details the benefits and use of co-occurrence and statistic positional concordance analysis and contains short examples for both techniques.

8.1.1 Co-Occurring Words: Finding and Interpreting Patterns of Semanticization and Cultural Attributions

Automated co-occurrence analysis is especially useful for studies of themes and semantic patterns throughout the breadth of the whole of a corpus. The co-occurrence of words can tell us something about the semantic qualities of a word (of course limited to its use within this specific corpus), e.g. its "semantic association" (Hoey 2007, 8), the meanings with which a word is associated; its "semantic prosody" (Louw 1993, 157; cf. Sinclair 1998: 15-6), i.e. does it show a tendency to co-occur with terms of positive, negative or neutral evaluation; or the "semantic preference" (Sinclair 1998, 15-6) of a term, resp. the semantic field the co-occurring words predominantly belong to (cf. Stubbs 2001, 65-6).[21] Through analyzing co-occurrences quantitatively across all novels of one city, the testing of qualitatively derived impressions, hypotheses and interpretations is possible. We can get a better grip on the semantic webs in which words are embedded, the discourse surrounding topics and themes, the meanings and cultural attributions which are ascribed to the whole of the city or each of its districts, and aspects, topics and themes attributed to them.[22] The following example (figure 2) shows the network graph of the 30 most frequent co-occurring nouns for the "Bahnhofsviertel" (district around the central station)

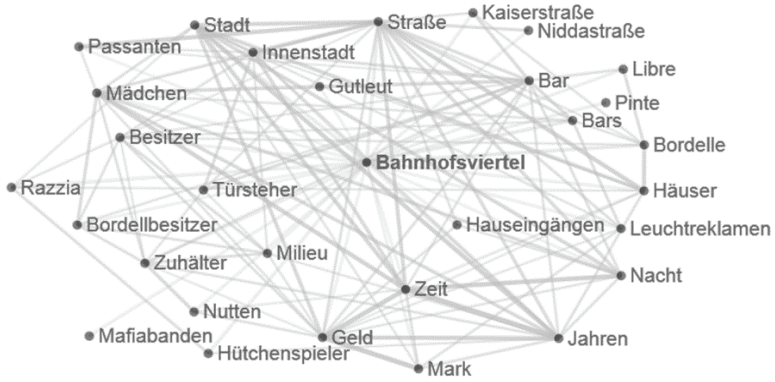
[20] This aspect can be further automatized through entering the respective information (author/book and device or word found) into a table of characteristics and processing it with a statistical program like SPSS.

[21] Semantic prosody and semantic preference are closely related and sometimes blurred concepts, see for example the discussion of this terms in Partington (2004) and Bednarek (2008).

[22] An exemplary and comparative analysis of the use of each city's name can be found in Rauscher et al. (2013, 66-8).

of Frankfurt on paragraph level (i.e. those nouns coming up repeatedly within the same paragraph as the designation of the district). During the initial exemplary close reading, we found the tendency to represent this area in close connection with its almost classical cliché as red light district, and wanted to find out if this holds true for the entire corpus. The search for nouns – displaying topics, themes, and people associated with this city district, its prevalent cultural ascriptions and associations – can provide a good overview of 'what is going on' in an area resp. what is frequently associated with it.

Figure 2: Top-30 Co-Occurring Nouns for "Bahnhofsviertel" in the Frankfurt Corpus



As the graph displays, the name of the district shows a semantic preference towards the sleazy world of pimps (Zuhälter) and prostitutes (Nutten). It is associated with the nightlife (Nacht (night), Bar, Bars, Pinte (pub)), its red light district forming a central topic for its literary staging (as can be seen with the co-occurrence of Bordell (brothel), Bordellbesitzer (brothel owner), Milieu, Mädchen (girls)), and money (Geld and Mark) playing a role, too. With the exception of "Passanten" (passersby), each mentioned group of people belongs to the sleazy side of this district or to the criminal scene that has been established there (Mafiabanden (mafia groups), Hütchenspieler (thimblerigger)) – making police action necessary as it seems (Razzia (raid)). Additionally, its bordering districts are mentioned (Innenstadt and Gutleut) as well as its most prominent streets (Kaiserstraße and Niddastraße). Even that the Kaiserstraße seems to be more prominent for pubs and brothels than the Niddastraße (the latter only showing connections to the search term and the anonymous "Straße" (street)) can be seen from the graph through its cross-connection with "Bar" and "Bordell". If we integrate the alternative designations for the same area (which can be found easily through the drop-down list function of the search field, displaying all words contained in the resp. corpus that start with

"Bahnhof-"), such as "Bahnhofsgebiet", we can confirm this analysis and add some more aspects and topics to it: more terms for pub (Kneipe, Kneipen) occur, as well as "Körperverletzung" (assault), but also more terms indicating police action (Einsatz (operation), Streife (patrol), Ermittlungen (investigations)). This confirms our impressions and findings from the previous steps of our analysis also on word-level and gives a very good overview of the frequently occurring topics, themes and cultural associations with this district in its literary staging.

8.1.2 Statistical Positional Concordances: Co-Occurring Words with a Twist

Concordance analysis is a technique for displaying and re-ordering sentences that are retrieved automatically with the help of a word search. In CoocViewer it is coupled with a form of automated co-occurrence analysis, so that the retrieved words are arranged automatically according to their frequency and significance of co-occurrence, and their position in relation to the search term. It widens the amount of co-text automatically analyzed for co-occurrences to 10 words to the left and right of the search term. Statistical positional concordances make comparison of semanticizations (and therefore meaning) of the same or similar terms in different corpora even more easy than the pure co-occurrence graphs, because the ordering aids the identification of patterns of similarity or contrast between cities. Principally, it fulfills the same functions resp. brings the same opportunities and advantages as the co-occurrence analysis, but enhances this instrument through the additional information contained in the graphs. The following example is a comparison between the Glasgow and Birmingham corpus for the top 30 co-occurrences with the word "streets" on sentence level (figure 3, Birmingham is presented in the upper, Glasgow in the lower graph). Both corpora show similar frequencies for the search term: it occurs 482 times in the Glasgow corpus and 293 times in that of Birmingham, but due to their differing sizes, in both corpora it has a parts-per-million value around 60 (62ppm for Glasgow, 60ppm for Birmingham). This time, we regard the adjectives, verbs and adverbs that surround this word within a range of -3 to +3 to investigate frequent actions on the streets as well as their recurring description.

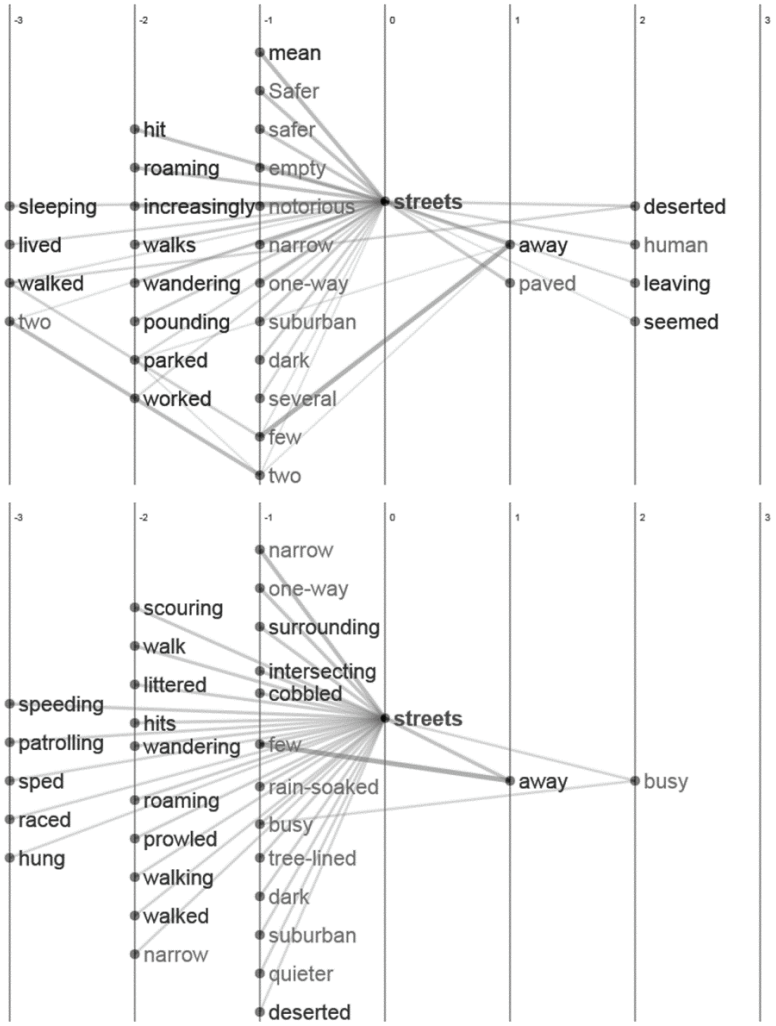
The graphs reveal a differing vocabulary surrounding the search term in both corpora, with a slight tendency in the Glasgow corpus of more "fixed" verbs (i.e. occurring on the same position across different authors) that precede it (especially on position -3), while Birmingham shows more frequent recurring ones on the positions +1 and +2 (like in "the streets were deserted"). For the position directly left to the node, where the words displayed directly alter the noun "streets", both cases show adjectives (and verbs) describing the quality or look of streets or their spatial relation ("surrounding" and "intersecting" for Glasgow). In the Glasgow corpus, the words "narrow" (most common with 10

co-occurrences at this position and three for the -2 position), "one-way", "rain-soaked", "busy", "tree-lined", "dark", "quieter", and "deserted" pose relatively usual ways for describing the quality of streets, while Birmingham also features adjectives of negative evaluation: "mean" (here displayed as verb due to its predominant use as such throughout the corpus) – the word that is highest ranked in the graph (10 co-occurrences at this position) with a distribution across 10 different novels and three authors – and "notorious". This negative semantic prosody was rather unexpected and stands opposite to our impressions from the qualitative reading: we anticipated that "streets" in Glasgow would (frequently) be described with terms carrying a negative evaluation or evoking some sense of fear or violence, since the descriptions that struck us as readers where those of "mean streets". But it seems that we overestimated this aspect, or that the differing authors use a much more varied language to evoke negative descriptions or allusions to danger. But, if we look at the graph again, why should "Safer" (or "safer"), paradoxically, be the second most common co-occurrence after "mean" in the case of Birmingham? A closer look at the references reveals that those instances of "Safer/safer", with a co-occurrence of three each, do not express a quality of the streets, but are due to one novel where the author describes the scene of a "Women Against Rape March", where the slogan "Safer streets…cage the beast" (Carter 2006, 124-6) is repeatedly chanted by the crowd. Therefore, we can interpret them first as author resp. even novel specific, and, second, as affirmation for the "mean" and "notorious" streets of Birmingham's literary staging. Besides these qualities of the streets, both corpora also display words indicating quantity: "several", "few" and "two" for Birmingham, only "few" for Glasgow (and in both cases, the triangular link between "few", "streets" and "away", spanning the positions -1, 0 and +1, shows the meaning of this subset through indicating that they occur together in this way). There seems to be no indication on sentence level for the frequent mentioning of the topic of 'traffic' in relation with the term "streets" in the case of Birmingham (as we would have somewhat suspected according to our previous analysis).

The verbs that co-occur frequently with "streets" are different in both corpora, too. Especially those on the -3 position to the left of the node seem to hint at an interesting pattern. While in the Glasgow corpus, somebody is "speeding through the streets" (or "sped" through them, together five co-occurrences across five different texts and authors) or is "patrolling" them, Birmingham here shows the pattern of people "sleeping on the streets" resp. having "lived" on them for some time (thereby further contributing to the mean-streets impression). Also the verb "worked" (the least frequent on position -2 with only two co-occurrences within the same novel) can be seen as belonging to this 'seedy-streets' impression, being used as "worked the streets" (for street prostitution). Besides this, both corpora show (slightly) differing semanticizations when it comes to walking on the streets, with "roaming" being a bit more frequent in

the Birmingham corpus (with a relative frequency of 2.26ppm in contrast to Glasgow's 1.55ppm), but Glasgow showing more vocabulary richness. Another interesting difference that cannot directly be deduced from the graphs is the different use of the term "hit(s)" that can be observed through a comparative look at the references.

Figure 3: Statistical Positional Concordances in Birmingham (Upper) and Glasgow (Lower)



While in Birmingham news and detectives "hit the streets", this expression is used in Glasgow only for "smack" or heroin when it "hits the streets".

Although these brief exemplary analyses from our spatial dimension can only give a restricted insight into the mass of individual analysis that need to be carried out before we can evaluate and enhance our background hypothesis of the "intrinsic logic of cities", the data analyzed so far with this visual method strengthens our impression that there are indeed city-specific patterns and structures within their literary representations. These also comprise the micro-level of words used for the description and depiction of urban life and the cultural attributions towards urban space and places.

8.2 Re-Integrating the Close Perspective and Starting the Feedback Loop

It should have become clear by now in how far visualization as method can contribute to the investigation of literary representations across a high amount of literary texts on the micro-level of words. Nevertheless, it also should have become clear why distant reading and visualization alone are not sufficient: the absolute or relative frequency of one word alone can tell us nothing about its use, function or meaning in its context, and the references of the co-occurrences have to be checked and read to forestall premature evaluations and conclusions (as shown with the example of the "safer streets" in Birmingham). This particular and limited close reading can (and should) be enhanced and enlarged to a wider context than the simple sentence of co-occurrence, especially in those cases where the sentence alone does not reveal the meaning or seems to be ambiguous (e.g. because the sentence refers to a subject given in a previous one, or to come to terms with the rhetorical device of irony). The return to the QCA software and the document(s) where the respective reference(s) occurred makes the additional close reading of the wider context surrounding the co-occurrence easy and is one possibility for moving from close to distant and back again during the analyzing process. The results of the co-occurrences and concordance analysis and the corresponding segments of the texts can also be coded during this 'step back' to complete the analysis. Taking the other way round, aspects and terms occurring at a mid-level distance during the computationally aided qualitative content analysis can be quickly fed into the step of distant reading, giving an overview of their relative frequency to establish if the respective aspect plays a role across the whole of a corpus, or if it does so also in other corpora (or in differing proportions, etc.). During phase two, the process of analysis basically consists of the back and forth between content analysis and lexical analysis and between different degrees of context taken into account.

Another problem that can be addressed through such a feedback loop to close reading resp. the middle-distance of the QCA data base is the literariness

of the texts under investigation. A high variety in diction between the different authors for the same or similar aspects needs to be considered, as the relatively low frequencies of the exemplary concordance analysis indicate. Therefore, besides those references of frequent co-occurrences, also the non-frequent co-occurring words (i.e. those that do not come over the two-occurrence threshold) should be taken into account through investigating all references for the search term. Additionally, synonymies or alternative spellings of the search term should be analyzed, too, before those results are fed back into MaxQDA 10 and so on.

After analyzing the aspects that fall within the dimensions as developed during the first phase with the aid of different computer-programs as detailed during the previous sections, one last round of 'reading the book' should be included for those novels or authors that did not or only very seldom appear in the analysis so far, i.e. those with the lowest 'hit rate', to investigate these exceptions or variations of the patterns. This strengthens the overall results and exposes patterns, structures or variations that were possibly overlooked – which, after exploring them, can be fed back into the cyclical or iterative research process and then included into the final interpretation of the results.

9. Conclusion

The analyses carried out so far show that literary representations, at least those stemming from the genre of crime fiction, indeed can provide a rich and valuable set of data for different (interdisciplinary) questions about the characteristic and distinctiveness of cities regarding manifest contents as well as discursive aspects. Contemporary crime novels contain and detail lots of information about the specific spatial and social context formed by their respective setting. The investigation of the staging of spatial aspects, frequent topics, themes and aspects of city life can give a rough overview of 'what is going on' and what is talked or written about in the respective city. Their semanticization and related lexical patterns generate insights into cultural attributions and reveal differing discourses and differing discourse strategies deployed in narrating the respective city. It can answer questions about the recurring elements within literary representations that form a part of the cumulative texture of each city and are simultaneously based on this texture. The further triangulation of these results with data from the production and reception of the novels and with the results of the other projects investigating the possibility of an 'Eigenlogik' of cities will show in how far homologue patterns can be found in the fields of economic and city marketing practices and in the discourse analysis of problem discourses in newspaper articles.

The existing gap between qualitative and quantitative methods in literary studies that often prevents a pragmatic implementation of methods stemming

from both fields has been addressed through a connection of close and distant reading in a multi-phased and iterative research process, where the strengths of both approaches are connected through their consecutive as well as concurrent use. While the qualitative close reading can generate the general dimensions of interest and first hypotheses about the ways in which the distinctiveness of cities plays a role in their literary staging (and plays a great role in the actual evaluation or interpretation of automatically extracted passages and lexical patterns), the quantitative methods from a more sociologically oriented content analysis and the distant reading approach can help to test these hypotheses, to quantify the results and to explore patterns across the breadth of the whole of each corpus on the mid-level of topics and the micro-level of words. The mixing of qualitative and quantitative approaches for the analysis of potentially city-specific patterns of literary representation produces complementary perspectives on the same data and a connection of different levels of analysis. It also includes and connects different 'portions' of context (direct co-text, surrounding paragraphs of search terms or the wide and unlimited context of the novel as whole) into the analysis which are important to tackle some of the problems that arise from the literariness of the data.

Through the alternate use of different computer programs a maximum of texts can be analyzed and interpreted, much more than any traditional approach from literary studies could process. Additionally, the visual approach towards the analysis can reveal patterns that would otherwise be unrecognized and helps to simplify the application of corpus linguistic tools and concepts, thereby lowering the threshold that these accounts usually pose to lay-people in this field. The systematical re-integration of close reading as a method for interpretation into the visual account of distant reading can prevent premature interpretations, relating the graphs back to their origin within each individual novel.

The method we developed is suited to address questions from all relevant dimensions, not only the spatial, and also can be used to get more insights into genre conventions (e.g. through an exploration of the uses and semanticizations of more genre-related terms, such as detective, murder or corpse). It is not limited to the exploration and analysis of literary corpora or the distinctiveness of cities. It is applicable to all kinds of social or literary research that poses qualitative research questions to large quantities of textual data (e.g. newspaper or magazine articles) or that is interested in aspects of content analysis and rhetoric or discourse analysis at the same time.

References

Alewyn, Richard. 1998. Anatomie des Detektivromans. In *Der Kriminalroman. Poetik, Theorie, Geschichte*, ed. Jochen Vogt, 52-72. München: Fink.

Allison, Sahra, Ryan Heuser, Mathew Jockers, Franco Moretti, and Michael Witmore. 2011. Q*uantitative Formalism: an Experiment*. Stanford Literary Lab, Pamphlet 1 (January 2011) <http://litlab.stanford.edu/LiteraryLabPamphlet1.pdf> (accessed August 29, 2013).

Batuman, Elif. 2006. Adventures of a Man of Science: Moretti in California. Review of *Graphs, Maps, Trees: Abstract Models for Literary History,* by Franco Moretti. N+1, 3 (Fall 2006), n.p. <http://nplusonemag.com/adventures-man-science> (accessed August 29, 2013).

Beatie, Bruce A. 1979. Measurement and the Study of Literature. *Computers and the Humanities* 13: 185-94 <http://www.jstor.org/stable/30207255>.

Bednarek, Monika. 2008. Semantic Preference and Semantic Prosody re-examined. *Corpus Linguistics and Linguistic Theory* 4 (2): 119-39. doi: 10.1515/CLLT.2008.006.

Berking, Helmuth, and Martina Löw, eds. 2008. *Die Eigenlogik der Städte: Neue Wege für die Stadtforschung*. Frankfurt a.M., New York: Campus.

Berking, Helmuth. 2008. Städte lassen sich an ihrem Gang erkennen wie Menschen – Skizzen zur Erforschung der Stadt und der Städte. In *Die Eigenlogik der Städte: Neue Wege für die Stadtforschung,* ed. Helmuth Berking and Martina Löw, 15-32. Frankfurt a.M., New York: Campus.

Berking, Helmuth. 2012. The distinctiveness of Cities: Outline of a Research Programme. *Urban Research & Practice* 5 (3): 316-24. doi: 10.1080/17535069.2012.727549.

Biber, Douglas. 2011. Corpus linguistics and the study of Literature: Back to the Future? *Scientific Studies of Literature* 1 (1): 15-23. doi: 10.1075/ssol.1.1.02bib.

Burrows, John F. 1987. *Computation into Criticism: A Study of Jane Austen's Novels and an Experiment in Method*. Oxford: Clarendon.

Burrows, John F. 1992. Computers and the Study of Literature. In *Computers and Written Texts*, ed. Christopher S. Butler, 167-204. Oxford: Blackwell.

Burrows, John F. 2007. All the way through: Testing for Authorship in different Frequency Strata. *Literary and Linguistic Computing* 22 (1): 27-47. doi: 10.1093/llc/fqi067.

Börsenverein. 2011. *Buch & Markt. Belletristik*. <http://www.boersenverein.de/de/portal/Belletristik/189810> (accessed August 29, 2013).

Brummett, Barry. 2010. *Techniques of Close Reading*. Los Angeles: Sage.

Carlin, Andrew P. 2010. The Corpus Status of Literature in Teaching Sociology: Novels as 'Sociological Reconstruction'. *American Sociologist* 41 (3): 211-31. doi: 10.1007/s12108-010-9096-8.

Carter, Maureen. 2006. *Baby Love*. Chesterfield: Crème de la Crime.

Clement, Tanya E. 2008. 'A thing not Beginning and not Ending': Using Digital Tools to Distant-Read Getrude Stein's *The Making of Americans. Literary and Linguistic Computing* 23 (3): 361-81. doi: 10.1093/llc/fqn020.

Corns, Thomas N. 1991. Computers in the Humanities: Methods and Applications in the Study of English Literature. *Literary and Linguistic Computing* 6 (2): 127-30. doi: 10.1093/llc/6.2.127.

Craig, Hugh. 1999. Jonsonian chronology and the Styles of a Tale of a Tub. In *Re-Presenting Ben Jonson: Text, History, Performance,* ed. Martin Butler, 210-32. Houndmills: Macmillan.

Craig, Hugh. 2004. Stylistic analysis and authorship studies. In *A Companion to Digital Humanities,* ed. Susan Schreibman, Ray Siemens and John Unsworth, 273-88. Oxford: Blackwell.

Creswell, John W. 2009. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches,* 3rd ed. Los Angeles a.o.: Sage.

Dalton, John. 2002. *The City Trap*. Birmingham: Tindal Street Press.

Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19 (1): 61-74.

Evers, Stuart. 2009. Why Scene is as Crucial as Crime in Detective Fiction. *The Guardian*, March 3. <http://www.theguardian.com/books/booksblog/2009/feb/25/crime-fiction-setting-padura> (accessed August 29, 2013).

Fischer-Starke, Bettina. 2010. *Corpus linguistics in literary analysis: Jane Austen and her contemporaries*. London: Continuum.

Franzosi, Roberto. 2010. *Quantitative Narrative Analysis*. Los Angeles: Sage.

Früh, Werner. 2011. *Inhaltsanalyse: Theorie und Praxis*, 7th revised ed. Konstanz, München: UVK Verlagsgesellschaft mbH.

Fühlau, Ingunde. 1982. Die Sprachlosigkeit der Inhaltsanalyse: Linguistische Bemerkungen zu einer sozialwissenschaftlichen Analyse. *Kodikas/Code Supplement* 14. Tübingen: Gunter Narr.

Genette, Gérard. 1994. *Die Erzählung,* 3rd ed. München: W. Fink.

George, Alexander L., and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, Mass: Belfer Center for Science and International Affairs.

Hayles, N. Katherine. 2013. Combining Close and Distant Reading: Jonathan Safran Foer's Tree of Codes and the Aesthetic of Bookishness. *Publications of the Modern Language Associations of America (PMLA)* 128 (1): 226-31. doi: 10.1632/pmla.2013.128.1.226.

Hoey, Michael. 2007. Lexical priming and literary creativity. In *Text, Discourse and Corpora: Theory and Analysis,* ed. Michael Hoey, Michaela Mahlberg, Michael Stubbs and Wolfgang Teubert, 31-56. London: Continuum.

Horsley, Lee. 2005. *Twentieth Century Crime Fiction*. Oxford: Oxford University Press.

Hoover, David L. 2001. Statistical Stylistics and Authorship Attribution: an Empirical Investigation. *Literary and Linguistic Computing* 16 (4): 421-44. doi: 10.1093/llc/16.4.421.

Hoover, David L. 2002. Frequent Word Sequences and Statistical Stylistics. *Literary and Linguistic Computing* 17 (2): 157-80. doi: 10.1093/llc/17.2.157.

Hoover, David L. 2008. Quantitative analysis and literary studies. In *A Companion to Digital Literary Studies,* ed. Susan Schreibman and Ray Siemens, chap. 28. Oxford: Blackwell <http://www.digitalhumanities.org/companionDLS/>.

Hori, Masahiro. 2004. *Investigating Dickens' Style: A Collocational Analysis.* Basingstoke: Palgrave Macmillan.

Hunston, Susan. 2006. Corpus Linguistics. In *Encyclopedia of Language and Linguistics*, 2nd ed., ed. Keith Brown, 234-48. Oxford: Elsevier.

Kelle, Udo. 2007. Integration qualitativer und quantitativer Methoden. In *Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis*, 2nd ed., ed. Udo Kuckartz, Thomas Ebert, Stefan Rädiker and Claus Stefer, 50-64. Wiesbaden: VS Verlag für Sozialwissenschaften.

Khadem, Amir. 2012. Annexing the Unread: a Close Reading of Distant Reading. *Neohelicon* 39: 409-21. doi: 10.1007/s11059-012-0152-y.
Kuckartz, Udo, Thomas Ebert, Stefan Rädiker, and Claus Stefer, eds. 2007. *Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis.* Wiesbaden: VS Verlag für Sozialwissenschaften.
Kuckartz, Udo. 2010. *Einführung in die computergestützte Analyse qualitativer Daten.* Wiesbaden: VS Verlag für Sozialwissenschaften.
Liddle, Dallas. 2012. Reflections on 20,000 Victorian Newspapers: 'Distant Reading' *The Times* using *The Times Digital Archive*. *Journal of Victorian Culture* 17 (2): 230-37. doi: 10.1080/13555502.2012.683151.
Louw, William Ernest. 1993. Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies. In *Text and Technology: In Honour of John Sinclair*, ed. Mona Baker, Gill Francis and Elena Tognini-Bonelli, 157-76. Amsterdam and Philadelphia: John Benjamins.
Löw, Martina. 2008a. *Soziologie der Städte*. Frankfurt a.M.: Suhrkamp.
Löw, Martina. 2008b. Eigenlogische Strukturen – Differenzen zwischen Städten als konzeptuelle Herausforderung. In *Die Eigenlogik der Städte: Neue Wege für die Stadtforschung,* ed. Helmuth Berking and Martina Löw, 33-54. Frankfurt a.M., New York: Campus.
Löw, Martina. 2012. The Intrinsic Logic of Cities: Towards a New Theory on Urbanism. *Urban Research & Practice* 5 (3): 303-15. doi: 10.1080/17535069.2012.727545.
Löw, Martina. 2013. The City as Experiential Space: The Production of Shared Meaning. *International Journal of Urban and Regional Research* 37 (3): 894-908. doi: 10.1111/1468-2427.12022.
Mahlberg, Michaela. 2007. Corpus stylistics: Bridging the Gap between Linguistic and Literary Studies. In *Text, Doscourse and Corpora: Theory and Analysis*, ed. Michael Hoey, Michaela Mahlberg, Michael Stubbs and Wolfgang Teubert, 217-46. London: Continuum.
McKenna, C., F. Wayne, and Antonia Alexis. 2001. The Statistical Analysis of Style: Reflections on Form, Meaning, and Ideology in the 'Nausicaa' Episode of Ulysses. *Literary and Linguistic Computing* 16 (4): 353-73. doi: 10.1093/llc/16.4.353.
Membery, York. 2013. Top 10 Criminal Masterminds: From Patricia Cornwell to David Baldacci, the world's most successful living crime writers revealed. *Daily Mail,* May 25 <http://www.dailymail.co.uk/home/event/article-2325521/Patricia-Cornwell-David-Baldacci-The-worlds-successful-living-crime-writers-revealed.html> (accessed August 29, 2013).
Mitrić, Ana. 2007. Jane Austen and Civility: A Distant Reading. *Persuasions (The Jane Austen Journal)* 29: 194-208.
Moretti, Franco. 2000a. Conjectures on World Literature. *New Left Review,* 1 (January, Feburary): 54-68.
Moretti, Franco. 2000b. The Slaughterhouse of Literature. *Modern Language Quaterly* 61 (1): 207-27. doi: 10.1215/00267929-61-1-207.
Moretti, Franco. 2007. *Graphs, Maps, Trees: Abstract Models for Literary History.* London, New York: Verso.
Moretti, Franco. 2009. Style, Inc.: Reflections on seven thousand Titles (British Novels, 1740-1850). *Critical Inquiry* 36 (1): 134-58. doi: 10.1086/606125.

Moretti, Franco. 2011. *Network Theory, Plot Analysis*. Stanford Literary Lab, Pamphlet 2 (May 2011) <http://litlab.stanford.edu/LiteraryLabPamphlet2.pdf> (accessed August 29, 2013).
Moretti, Franco. 2013. *Distant Reading.* London*,* New York: Verso.
Partington, Alan. 2004. 'Utterly content in each other's company': Semantic prosody and semantic preference. *International Journal of Corpus Linguistics* 9 (1): 131-56 doi: 10.1075/ijcl.9.1.07par.
Potter, Rosanne G. 1988. Literary Criticism and Literary Computing: The difficulties of Synthesis. *Computers and Humanities* 22: 91-7 <http://www.jstor.org/stable/30200105>.
Prendergast, Christoph. 2005. Evolution and Literary History. *New Left Review* 34 (July-August): 40-62.
Rapaport, Herman. 2011. *The Literary Theory Toolkit: A Compendium of Concepts and Methods*. Malden, Mass: Wiley-Blackwell. doi: 10.1002/9781444395693.
Rauscher, Janneke, Leonard Swiezinski, Martin Riedl, and Chris Biemann. 2013. Exploring Cities in Crime: Significant Concordance and Co-occurence in Quantitative Literary Analysis. *Proceedings of the Second Workshop on Computational Linguistics for Literature*, *Atlanta, Georgia, June 14, 2013*, 61-71 <http://aclweb.org/anthology/W/W13/W13-1409.pdf> (accessed September 3, 2013).
Rommel, Thomas. 2004. Literary Studies. In *A Companion to DigitalHumanities*, ed. Susan Schreibman, Ray Siemens and John Unsworth, ch. 8. Oxford: Blackwell <http://www.digitalhumanities.org/companion/>.
Schmidt, Jochen. 1989. *Gangster, Opfer, Detektive: Eine Typengeschichte des Kriminalromans*. Frankfurt a.M., Berlin: Ullstein.
Serlen, Rachel. 2010. The Distant Future? Reading Franco Moretti. *Literature Compass* 7 (3): 214-25. doi: 10.1111/j.1741-4113.2009.00669.x.
Sinclair, John. 1998. The Lexical Item. In *Contrastive Lexical Semantics*, ed. Edda Weigand, 1-24. Amsterdam, Philadelphia: John Benjamins.
Stolz, Matthias. 2013. Deutschlandkarte: Regionalkrimis. *ZEITMagazin*, January 16 <http://www.zeit.de/2013/03/Deutschlandkarte-Regionalkrimis> (accessed August 29, 2013).
Stone, Phillip J., Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge, Mass: MIT Press.
Stubbs, Michael. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
Stubbs, Michael. 2005. Conrad in the Computer: Examples of Quantitative Stylistic Methods. *Language and Literature* 14 (1): 5-24. doi: 10.1177/0963947005048 873.
Suerbaum, Ulrich. 1984. *Krimi: Eine Analyse der Gattung*. Stuttgart: Reclam.
Suttles, Gerald D. 1984. The Cumulative Texture of Local Urban Culture. *American Journal of Sociology* 90 (2): 283-304.
Van Peer, Willie. 1989. Quantitative Studies of Literature: A Critique and an Outlook. *Computers and the Humanities* 23: 301-7. doi: 10.1007/BF02176635.